

What is Claimed:

1. A computer-implemented method of ranking the relevancy of a collection of hypertext pages to a keyword-based query, comprising:
calculating an intrinsic rank of a page;
calculating an extrinsic rank of the page; and
calculating the rank of the page by combining the intrinsic rank and the extrinsic rank.
2. The method of claim 1, wherein the intrinsic rank is a function of the content score and the page weight of the page.
3. The method of claim 2, wherein the content score is a function of the frequency, location, and/or font size of a keyword in the page.
4. The method of claim 2, wherein the page weight is defined as the probability of a user visiting the page when traveling in the collection of hypertext pages in a random fashion.
5. The method of claim 2, wherein the page weight is obtained as the sum of the product of a link weight of each inbound link to the page and the page weight of the originating page.
6. The method of claim 2, wherein the page weight is computed by the following steps of:
constructing a connectivity graph, which represents the collection of hypertext pages and the link structure between the pages;
adding a page weight reservoir with bi-directional links to and from each of the pages in the collection of hypertext pages; and
summing all of the products of each inbound link weight with the page weight of the originating page providing the inbound link.

7. The method of claim 2, further comprising computing the page weights by the following steps of:

initializing a page weight vector to a constant;

constructing a connectivity graph representative of the link structure of the collection of pages;

computing an output page weight vector from the input page weight vector and the connectivity graph; and

comparing the output page weight vector with the input page weight vector for convergence, and if convergence is reached, writing the output page weight vector in a page weight database, and if not, mixing the input and output page weight vectors to generate a new input page weight vector and repeating until convergence is reached.

8. The method of claim 5, wherein the link weight is defined as the probability of a user randomly choosing the link to visit other pages when traveling in the collection of hypertext pages.

9. The method of claim 5, wherein the link weight of the inbound links has a uniform value corresponding to the reciprocal of the total number of links outbound from an originating page.

10. The method of claim 5, wherein the link weight has a variable value, which depends on the number of outbound links, the offset of the link, the size of the paragraph where the link is located, and/or whether the link is an external or internal link.

11. The method of claim 1, wherein the extrinsic rank is a function of the anchor weight and the page weight of the pages providing inbound links to the page.

12. The method of claim 1, wherein the extrinsic rank is obtained by summing the products of the anchor weight and the page weight of the originating page providing each inbound link.

13. The method of claim 11, wherein the anchor weight is a function of the inbound link weights and the keyword being present in the anchor text, in the vicinity of the anchor text, or in text related to the topic of the anchor text.

14. The method of claim 11, wherein the page weight is defined as the probability of a user randomly visiting a page in the collection of hypertext pages.

15. The method of claim 11, wherein the page weight is obtained by summing the products of the link weight of each inbound link to the page and the page weight of the originating page providing the inbound links.

16. The method of claim 11, wherein the page weight is computed by the following steps of:

- constructing a connectivity graph, which represents the collection of hypertext pages and the link structure between the pages;
- adding a page weight reservoir with bi-directional links to and from each of the pages in the collection of hypertext pages; and
- summing all of the products of each inbound link weight with the page weight of the originating page providing the inbound link.

17. The method of claim 11, further comprising computing the page weights by the following steps of:

- initializing a page weight vector to a constant;
- constructing a connectivity graph representative of the link structure of the collection of pages;
- computing an output page weight vector from the input page weight vector and the connectivity graph; and

comparing the output page weight vector with the input page weight vector for convergence, and if convergence is reached, writing the output page weight vector in a page weight database, and if not, mixing the input and output page weight vectors to generate a new input page weight vector and repeating until convergence is reached.

18. The method of claim 15, wherein the link weight is defined as the probability of a user randomly choosing the link to visit other pages when traveling in the collection of hypertext pages.

19. The method of claim 15, wherein the link weight of the inbound links has a uniform value corresponding to the reciprocal of the total number of links outbound from an originating page.

20. The method of claim 15, wherein the link weight has a variable value, which depends on the number of outbound links, the offset of the link, the size of the paragraph where the link is located, and/or whether the link is an external or internal link.

21. The method of claim 1, wherein the collection of hypertext pages is fetched from the Web.

22. A computer-implemented method of ranking a collection of hypertext pages, comprising:

- calculating the intrinsic rank of a page for a multi-keyword query;
- calculating the extrinsic rank of the page for the multi-keyword query; and
- calculating the rank of the page in the collection of hypertext pages by combining the intrinsic rank and the extrinsic rank.

23. The method of claim 22, wherein the intrinsic rank is a function of content score and the page weight.

24. The method of claim 23, wherein the content score is a function of the proximity value of the multi-keywords and of the frequency, location, and/or font size of the multi-keywords in the page.

25. The method of claim 22, wherein the extrinsic rank of the page is a function of the partial extrinsic ranks and the proximity value of the multi-keywords.

26. The method of claim 25, wherein partial extrinsic rank is a function of the anchor weight and the page weight of the pages with identical anchor text.

27. The method of claim 25, wherein partial extrinsic rank is computed by summing the products of the anchor weight and the page weight of the pages with identical anchor text.

28. A Web search engine, comprising:

- a Web page database;
- a crawler to fetch pages from the Web and store the pages in the Web page database;
- a link extractor to extract link information from the pages;
- a URL management system to assign an identification number to the URL of each page, and store the identification number and URL pairs in the Web page database and send new URLs to the crawler to be retrieved from the Web;
- anchor text and link database;
- an anchor text and link extractor to extract the anchor text and the link information from the pages and store in the anchor text and link database;
- indexed database;
- an indexer to parse keywords from the pages and store the keyword and URL identification pairs in the indexed database; and

a ranker to rank a page based on intrinsic rank and extrinsic rank of the page.

29. The Web search engine of claim 28, wherein the ranker determines the intrinsic rank from content information in the indexed database and the page weight computed from the link information in anchor text and link database, and the extrinsic rank from the anchor text information in the anchor text and link database and the computed page weight.

30. The Web search engine of claim 28, wherein the ranker determines the intrinsic rank of the page based on the content score and the page weight.

31. The Web search engine of claim 28, wherein the ranker determines the extrinsic rank of the page based on the anchor weight of each inbound link and the page weight of the originating page.

32. The Web search engine of claim 28, wherein the ranker determines the anchor weight based on the link weight and the keyword being present in the anchor text or related text.

33. The Web search engine of claim 28, wherein the ranker calculates the intrinsic rank and extrinsic rank of a page for a multi-keyword query, wherein the intrinsic rank is a function of content score and the page weight, the extrinsic rank of the page is a function of the partial extrinsic ranks and proximity values.

34. The Web search engine of claim 28, further comprising a page weight generator and a page weight database, computing page weights by initializing a page weight vector to a constant, constructing a connectivity graph representing the link structure of the fetched pages, computing an output page weight vector from the input page weight vector and the connectivity graph, and comparing the output page weight vector with the input page weight vector and if convergence

is reached, writing the output page weight vector in a page weight database, and if not, mixing the input and output page weight vectors to generate a new input page weight vector and repeating until convergence is reached.

35. A computer system for ranking search results from a query on a collection of hypertext pages, comprising:

- a crawler to fetch pages from the collection of hypertext pages;
- a link extractor to extract page locator information from the fetched pages;
- a page locator management system for storing and retrieving the page locator information;
- a page database to store the pages;
- an indexer to parse keywords from the pages and store the keyword page locator pairs in the indexed database;
- an anchor text and link extractor to extract the anchor text and link structures from the pages;
- an anchor text and link database, wherein the anchor text and link extractor writes the anchor text and link structures into the anchor text and link database; and
- a ranker to assign a rank value to a page based on intrinsic and extrinsic rank.

36. The system of claim 35, wherein the ranker assigns an intrinsic rank to the page based on a combination of content score and page weight.

37. The system of claim 35, wherein the ranker assigns the content score to the page for a keyword based on a combination of location, frequency, and/or font size of the keyword in the page.

38. The system of claim 35, wherein the ranker assigns a page weight to the page as the probability of a searcher visiting the page when traveling in the collection of hypertext pages in a random fashion.

39. The system of claim 35, wherein the ranker assigns a uniform value corresponding to the reciprocal of the total number of links outbound from an originating page to link weight.

40. The system of claim 35, wherein the ranker assigns link weight based on location of the link.

41. The system of claim 35, wherein the ranker assigns an extrinsic rank to the page for a given keyword as a combination of anchor weight of the links from other pages and the page weight of referring pages.

42. The system of claim 35, wherein the ranker assigns a rank value to a page for a multi-keyword query as a combination of intrinsic rank and extrinsic rank for the multi-keyword.

43. The system of claim 35, wherein the ranker assigns an intrinsic rank to a page for a multi-keyword query as a combination of content score and page weight.

44. The system of claim 35, wherein the ranker assigns a content score to a page for a multi-keyword query as a combination of content score based on intersection of the given keywords and proximity value.

45. The system of claim 35, wherein the ranker assigns a partial extrinsic rank for each variation of identical anchor text.

46. The system of claim 35, wherein the ranker assigns a extrinsic rank to a page for a multi-keyword query as a combination of partial extrinsic rank of identical anchor text and proximity values in each anchor text.

47. The system of claim 35, wherein the ranker obtains a link connectivity graph of the pages.

47. The system of claim 35, wherein the ranker obtains the rank values from the link connectivity graph.

48. The system of claim 35, wherein the ranker calculates the page weight by iterative numerical procedure.

49. The system of claim 35, wherein the ranker accelerates the convergence of the iterative numerical procedure in obtaining connectivity rank scores.

50. The system of claim 35, wherein the ranker calculates rank values by dividing the pages into distinct number of groups.

51. The system of claim 35, further comprising a rate controller to control the rate of request for page retrieval.

52. The system of claim 35, wherein the Web page database stores the pages in a fixed record large enough to contain a predetermined percentage of all of the pages, wherein if the page is smaller, the fixed record has some empty space, and if the page is larger, the Web page database stores as much of the page as possible in the fixed record and the rest in a record file.